PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:	WO 93/04200
C12Q 1/68	A1	(43) International Publication Date:	4 March 1993 (04.03.93)

(21) International Application Number: PCT/US92/06701

(22) International Filing Date: 10 August 1992 (10.08.92)

(30) Priority data:
743,606 12 August 1991 (12.08.91) US

(71) Applicant: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West 7th Street, Austin, TX 78701 (US).

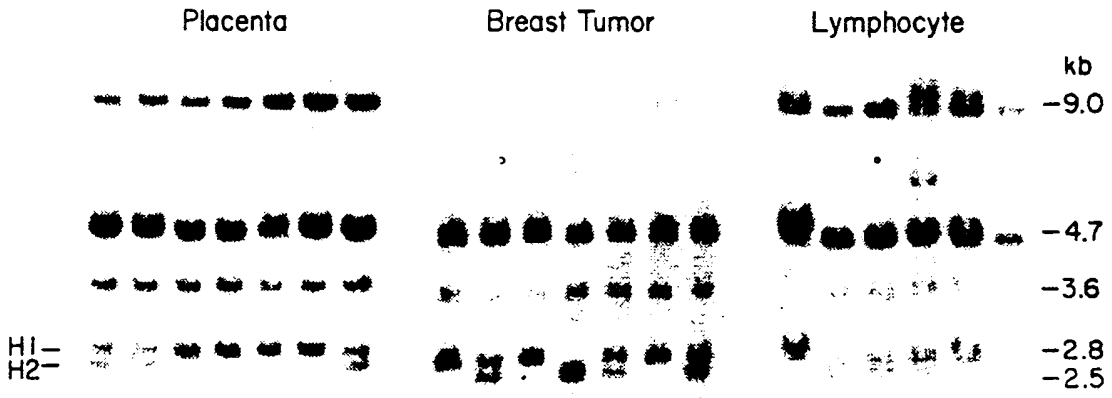
(72) Inventors: McGUIRE, William, L. ; 28433 Dalcin, San Antonio, TX 78260 (US). FUQUA, Suzanne, A. ; 2643 Inwood View, San Antonio, TX 78248 (US).

(74) Agent: MAYFIELD, Denise, L.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CS, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, PL, RO, RU, SD, SE, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.

(54) Title: METHODS AND COMPOSITIONS FOR THE DETECTION AND PREDICTION OF BREAST CANCER



(57) Abstract

A nucleic acid *Hin*dIII RFLP characteristic of human breast cancer or susceptiblilty to human breast cancer is presented. Restriction fragment length polymorphisms (RFLPs) are detectable using a restriction endonuclease capable of providing *Hin*dII DNA restriction fragments. The most frequent *Hin*dIII allelic profile of a person with breast cancer or susceptible to breast cancer is determined to be H2/H2 homozygous or to be absent an H1 allele. The *Hin*dIII RFLP may be used to detect or predict persons at risk of (or susceptible to) breast cancer through analysis of a wide variety of tissues, including breast tissues, tissue adjacent to a breast tumor, tissue and blood cells. A nucleic acid segment, particularly of 1.23 kb or less, and suitable for use as a probe for identifying specific DNA regions of the RFLP polymorphism, is also disclosed and constitutes part of a diagnostics kit for the identification of patients with or at risk of breast cancer.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MN	Mongolia
AU	Australia	FR	France	MR	Mauritania
BB	Barbados	GA	Gabon	MW	Malawi
BE	Belgium	GB	United Kingdom	NL	Netherlands
BF	Burkina Faso	GN	Guinea	NO	Norway
BG	Bulgaria	GR	Greece	NZ	New Zealand
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	PT	Portugal
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland			SE	Sweden
CI	Côte d'Ivoire	KR	Republic of Korea	SK	Slovak Republic
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CS	Czechoslovakia	LK	Sri Lanka	SU	Soviet Union
CZ	Czech Republic	LU	Luxembourg	TD	Chad
DE	Germany	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America

DESCRIPTION

METHODS AND COMPOSITIONS FOR THE DETECTION AND PREDICTION OF BREAST CANCER

The United States Government may own rights in the present invention as research relevant to the development thereof was supported by National Institutes of Health (NIAID) Grant No. CA30195

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to methods for detecting, as well as predicting, persons at risk of developing breast cancer. More specifically, the present invention provides a method for characterizing a person as having breast cancer, or as at risk of developing breast cancer, on the basis of a particular allelic profile manifest in persons with a particular nucleic acid genetic abnormality. Such a genetic abnormality is detectable in a *Hin*dIII RFLP. The RFLP of the present invention is further defined as being in close proximity or within to the progesterone receptor gene, most specifically on the long arm of the chromosome 11.

In particular embodiments, the RFLP of the present invention is identified in *Hin*dIII enzyme generated DNA fragments employing particularly defined nucleic acid segments suitable for use as DNA probes. The inventors have discovered that a certain percentage of persons whom have breast cancer or whom are characterized as "at risk" of developing this condition are found to be homozygous for the H2 allele (H2/H2) or are absent the H1 allele. Thus, it is either the presence of both H2 alleles, or the absence of the H1 allele, that serves to identify

persons susceptible to breast cancer or persons who already have breast cancer. As such, the present invention also relates to the field of diagnostic tests and kits for breast cancer.

2. Description of the Related Art

The steroid hormones, estradiol and progesterone, have profound effects upon the differentiation and proliferation of a number of target tissues, such as uterus and breast, and their actions are mediated by specific, low abundance intracellular receptors. The receptors for these hormones are known to modulate the transcription of specific genes, and are members of a superfamily of ligand-inducible enhancer factors[1].

The PgR, (progesterone receptor) in addition to being a gene regulator, is itself regulated by estrogen through the estrogen receptor (ER), thereby adding to the complexity of hormonal regulation in target tissues. PgR content has been found useful for predicting disease-free survival in human breast cancer, and is routinely measured in the clinical setting[2,3]. Several groups have now cloned the gene for PgRs from several species, and have demonstrated that human PgR is encoded by a single gene residing on chromosome 11[4-8]. This localization is characterized by frequent alterations and loss of chromosome 11 sequences in human breast cancer[9-11].

Chromosomal alterations and/or deletions are aberrations which have been proposed as mechanisms underlying oncogenesis. Indeed, amplification of specific oncogenes, such as c-erbB-2 and int-2, or alterations to c-H-ras-1, c-myc, and c-myb, may all contribute to either the genesis or the progression of human breast cancer[12-17].

Progesterone and breast cancer have been linked since the concept of hormone-dependent cancers was first elaborated[27-28]. In regard to biological factors, the literature reveals a wide range of potential physiological and biochemical factors as well as oncogene linked mechanisms, which result in the development of breast cancer.

Restriction fragment length polymorphism (RFLP) offers a powerful molecular genetic tool for the direct analysis of the human genome to determine elements that signal predisposition to genetic diseases[29]. Thus, such a technique permits the detection of variation in the human genome, and for the construction of genetic linkage maps. The technique of RFLP has been used to obtain information necessary to create genetic probes (isolated from chromosome-specific phage libraries) which contain a specific DNA portion of human DNA of interest.[30] With this tool in hand, the analysis of human gene segments is possible.

For example, restriction fragment analysis has been employed to examine relatively large segments of the human PgR gene which resides on chromosome 11.[8] The absence of PgR in a breast tumor could be due to a molecular alteration specifically in the PgR gene, or a consequence of defective regulation of the PgR gene by ER, or even to the concentration of steroid ligands themselves. These theories are supported by observations of others which report that PgR gene expression is under the dual control of both estrogen and progesterone[18,19]. However, a specific method for genetically identifying or predicting patients at genetic risk for developing breast cancer employing a part specific RFLP and specific patient allelic profile has not yet been reported.

The definition of specific gene abnormality in the breast tissue of a patient would provide an important advance in the art of breast cancer diagnostic tests. Even more importantly, a method which detects a genetic abnormality specific for breast cancer would be of potential value in objectively identifying individuals who are genetically predisposed to the development of breast cancer.

The elucidation of a specifically defined polymorphism in the DNA of a patient which was characteristic of breast cancer and conditions of occult breast cancer, would provide a valuable genetic marker for detecting as well as predicting the predisposition of a person for breast cancer.

SUMMARY OF THE INVENTION

The present invention provides a powerful specific method for the identification of particularly defined allelic polymorphisms useful in the diagnosis of human breast cancer. The present invention also provides methods whereby persons at risk for developing breast cancer may also be identified, i.e. in the diagnosis of occult breast cancer. These methods may be used in the analysis of nucleic acid obtained from tumorous tissue, such as from a breast tumor biopsy, as well as from non-tumorous tissues, such as blood, tissues adjacent a malignant breast tumor mass, skin, hair, buccal smear and thymus tissue.

The present inventors have discovered that the technique of RFLP may be employed in a method to detect the presence of a specific polymorphism(s) which indicates the presence of or a predisposition to breast cancer. This polymorphism may potentially be present in

the DNA found in cells of many tissue types. It is postulated that the presence of the specific polymorphism described herein may also be involved in the mechanism whereby a particular oncogene(s) is activated during the malignant disease process, or in the alternative, be involved in the suppression of a tumor suppressor gene.

The present inventors have discerned the existence of particular chromosome alterations in DNA obtained from breast tumor tissue. This particular RFLP is identifiable using a nucleic acid segment having a sequence defined in Figure 1. Test sample nucleic acid obtained from a patient which is found to include sequences hybridizable to a sequence or part of a sequence which corresponds to the *Hin*dIII RFLP of breast cancer, as defined in Figure 1, or a fragment thereof, are identified as positive for breast cancer.

The particular *Hin*dIII RFLP for human breast cancer employed in the disclosed methods and kits are further defined as being located in close proximity to or within the progesterone receptor (PgR) gene. The PgR gene is located at chromosome 11.

In one embodiment of the present invention, a method employing a specific nucleic acid segment characteristic of human breast cancer is defined. The particular nucleic acid segment thus constitutes a genetic marker detectable as a RFLP (restriction fragment length polymorphism) comprising a *Hin*dIII polymorphism. The described genetic marker of the present invention includes a *Hin*dIII polymorphism further defined by a particular allelic profile, being either an absence of an H1 allele or an H2/H2 homozygous allelic profile condition. Thus, persons having an H1/H2 allelic profile or the absence of an H2 allele (i.e., an H1/H1 allelic profile) would be identifiable as not having the

particular *Hin*dIII polymorphism characteristic of human breast cancer or susceptibility thereto, according to the claimed method.

In preferred embodiments, the RFLP of the present invention may be defined in terms of hybridization probe sequences that will bind to, and therefore identify, the RFLP. In one aspect, therefore, the H1 or H2 is defined as a nucleic acid segment that will hybridize specifically with a nucleic acid segment prepared with sequence characteristics of an H1 or H2 polymorphism. Nucleic acid segments of about 1.23 kb or less are preferred because of enhanced specificity and decreased monetary expense for the production of a shorter nucleic acid segment. More specifically, it is generally recognized that the longer a particular "identifying" (i.e. "probe") nucleic acid sequence is, the greater the possibility that hybridization will occur between parts of the "identifying" sequence unrelated to the particular genetic disorder of interest. Thus, by employing smaller nucleic acid segments as an "identifying" nucleic acid sequence, the inventors provide a diagnostic test having enhanced specificity for human breast cancer and having a decreased probability of reporting a "false positive" result for breast cancer or susceptibility thereto. An example of such a nucleic acid segment is the 1.23 kb nucleic acid sequence set forth in Figure 1. Even more preferred, however, will be shorter sequences that maintain specificity for both the H1 and H2 alleles. For example, such may be defined in the inventors prophetic 0.48 kb nucleic acid segment sequence set forth in figure 2.

In another embodiment of the present invention, a method for diagnosing breast cancer or susceptibility to breast cancer in a patient is provided. In a most preferred embodiment, the method comprises obtaining a

tissue or blood specimen from a patient, processing the specimen to obtain DNA, subjecting the DNA to a restriction enzyme capable of defining regions of a *Hin*dIII polymorphism, so as to provide DNA restriction fragments, and diagnosing a patient homozygous for the H2 allele or absent a H1 allele as having breast cancer or as being susceptible to breast cancer. More specifically, the DNA of the patient used in the described method is genomic DNA.

The DNA restriction fragments as prepared in the process of the claimed method are separated by length to provide isolated DNA restriction fragments. The isolated DNA restriction fragments so obtained are then most preferably probed to locate segments of a *Hin*dIII RFLP therein, said segments of the RFLP being hybridizable to a nucleic acid segment having an H1 or H2 allele. The inventors have found the *Hin*dIII RFLPs to reside at those patient DNA fragments which have a length of about 2.8 kb (H1 allele) and about 2.5 kb (H2 allele).

Even more specifically, the isolated DNA restriction fragments are probed with a nucleic acid segment suitable for use as a probe and having a sequence defined in Figure 1 or a fragment thereof. Even more preferably, a prophetic embodiment of the method employs a nucleic acid segment suitable for use as a probe and having a sequence as defined in Figure 2 or a fragment thereof.

In a particularly preferred embodiment of the claimed method, the DNA restriction fragments are separated by the process of electrophoresis. The inventors propose to discern smaller nucleic acid fragments which include even more narrowly defined nucleic acid sequences characteristic of the H1 and H2 allele. In such event, the nucleic acid fragments of a patient DNA sample need not first be separated by

electrophoresis. Instead, the presence of the H1 and the H2 alleles would be determined, and the allelic profile of the patient discerned and classified as either characteristic of breast cancer or susceptibility thereto, or not. Thus, invariant PgR bands, which appear at 3.6, 4.7 and 9.0 kb on the Southern Blot of Figure 3, need not be considered.

Typically, probes employed for detection purposes will be labeled to provide for their ready detection following hybridization. Of course, radioactive labels such as ^{32}P are generally the most sensitive for identification purposes, through, e.g., autoradiography. However, radioactive labels suffer from disadvantages due to the need for disposal of radioactive wastes and the short half-life of, e.g., ^{32}P. Therefore, it may be desirable to employ non-radioactive labels such as avidin/biotin or enzyme ligands such as alkaline phosphatase, horseradish peroxidase, etc., which may be detected through the use of colorimetric substrates.

While it is proposed that any of a variety of restriction enzymes may be used in the practice of the present invention, the restriction enzyme most particularly preferred is the *Hin*dIII restriction enzyme. A restriction enzyme capable of defining regions of a *Hin*dIII polymorphism may be used in conjunction with any and all of the described aspects of the present invention. Restriction enzymes found not to provide these requisites include EcoRI and Pvu II. In a most preferred embodiment of the claimed method, the restriction enzyme is *Hin*dIII restriction enzyme.

Turning now to a consideration of where the particular polymorphisms characteristic of breast cancer or susceptibility to breast cancer may be found, the present inventors have discovered that isolated DNA

restriction fragments obtained from the patient's DNA include the H1 allele to be located in a restriction fragment which has a length of about 2.8 kb. The H2 allele is to be found in a DNA restriction fragment which has a length of about 2.5 kb. The H1 allele or the H2 allele may be located according to the present methods by probing the isolated DNA restriction fragments with a nucleic acid segment having the sequence defined in Figure 1 or a fragment thereof. In one particularly preferred embodiment of the claimed method, the presence of an H2 allele or an H1 allele is identifiable in a DNA restriction fragment which hybridizes with a probe having a sequence defined in Figure 1, or a fragment thereof. In an even more particularly preferred prophetic embodiment of the claimed method, the presence of an H2 allele or an H1 allele may be identified in a DNA restriction fragment which hybridizes with a nucleic acid a sequence defined in Figure 2 or a fragment thereof.

A nucleic acid probe, preferably DNA, having a sequence as defined in Figure 1 may be prepared as an AccI/BamHI generated fragment of a PgR cDNA. This particular embodiment of the probe is a 1.23 kb DNA fragment of the PgR cDNA.

In a more narrowly defined embodiment of the present invention, a method of detecting a *Hin*dIII RFLP in DNA is provided. In this method, the *Hin*dIII RFLP identifies a polymorphism characteristic of breast cancer or susceptibility thereto. The method comprises treating the DNA sample with a restriction enzyme capable of producing a DNA restriction fragment having a *Hin*dIII polymorphism, to produce DNA restriction fragments, probing the DNA restriction fragments with a nucleic acid segment capable of identifying a *Hin*dIII RFLP, and identifying a *Hin*dIII RFLP in a segment of the separated DNA restriction fragment which hybridizes with the

nucleic acid segment or a fragment thereof. In the described method, the polymorphism for breast cancer or susceptibility to breast cancer is a deletion of an H1 allele. Alternatively, the polymorphism for breast cancer or susceptibility to breast cancer discernible with the described method is a H2/H2 allelic profile. In still another embodiment of the claimed method, the particular allelic profile characteristic of breast cancer or susceptibility thereto is defined in an H1 PgR gene-free allelic profile or in an H2/H2 PgR gene allelic profile.

In still another embodiment of the present invention, a nucleic acid segment suitable for use as probe and capable of identifying a *Hin*dIII RFLP characteristic of human breast cancer or susceptibility thereto, is provided. The particular nucleic acid segment most preferably has 1.23 kb or less. The RFLP identifiable using the described nucleic acid segment in turn identifies an H2/H2 allelic condition or identifies the absence of an H1 allele in the allelic profile of the patient. In a most particularly preferred embodiment of the described nucleic acid segment, the nucleic acid segment has a sequence as defined in Figure 1 or a fragment thereof. In a prophetic embodiment of the nucleic acid segment, the segment is defined as having a sequence as defined in Figure 2, or a fragment thereof. Even more specifically this particular prophetic embodiment of the nucleic acid fragment has a length of about 0.48 kb.

In still further embodiments, the present invention is directed to kits for the prediction of breast cancer or breast cancer susceptibility in a patient. Kits of the present invention may be defined generally as including a hybridization probe capable of hybridizing to an H1 and H2 RFLP, wherein the probe is comprised in a

suitable container, such as a test tube or vial. The probe, whether it be an RNA or DNA segment, will also preferably be suitably aliquoted to render it ready for use with little or no pre-experimentation. Furthermore, the probe container will generally be packaged in a larger container or box for easy transportation, shipping, etc.

As discussed above, the nucleic acid probe will generally comprise a segment of about 1.23 kb or less that will nevertheless faithfully hybridize, and therefore identify, H1 and H2 alleles. Probes of this size are preferred in that they provide greater diagnostic specificity for the diagnosis of breast cancer and are economical to prepare. For this reason, nucleic acid probes that comprise the sequence of figure 1, or a fragment comprising a hybridizable portion of the sequence, will be particularly preferred. As used herein, the term "hybridizable" portion, is intended to refer to nucleic acid segments that are long enough to form specific hybrids with the H1 or H2 allele. In a prophetic embodiment, the nucleic acid probe comprises that sequence of Figure 2.

The size of the particular nucleic acid fragment employed in the described methods and kits is not to be limited to those of 1.23 kb or less in size.

Virtually any size nucleic acid segment or fragment thereof which includes a sequence characteristic of the *Hind* III RFLP described herein, and being hybridizable at least in part to the nucleic acid segment defined in Figure 1 may be employed in the described methods.

In still further embodiments, kits of the present invention will include a means for detecting hybridization between the probe and an H1 or H2 RFLP,

typically a label located either on the probe, or contained on a separate molecule that can be made to bind specifically to the probe (such as a second nucleic acid probe, and avidin/biotin binding pair, etc.). Preferred labels comprise an enzyme or radioactive label.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: SEQUENCE OF THE 1.23 kb NUCLEIC ACID SEGMENT SUITABLE FOR USE AS A PROBE USED TO IDENTIFY THE BREAST CANCER POLYMORPHISM PRESENT IN A *Hind*III RFLP.

Figure 2: SEQUENCE OF A 0.48 kb NUCLEIC ACID SEGMENT PROPOSED AS SUITABLE FOR USE AS A PROBE FOR THE PROPOSED IDENTIFICATION OF A BREAST CANCER POLYMORPHISM PRESENT IN A *Hind*III RFLP.

Figure 3: SOUTHERN HYBRIDIZATION ANALYSIS OF DNA FROM BREAST TUMORS, PLACENTAS AND NORMAL LYMPHOCYTES DIGESTED WITH *Hind*III AND PROBED WITH A NUCLEIC ACID SEGMENT OF 1.23 kb CLONE. The presented *Hind*III RFLP provides a representative Southern blot of genomic DNA from human placenta, breast tumor and peripheral blot lymphocytes digested with the restriction enzyme *Hind*III. The two *Hind*III alleles H1 (about 2.8 kb) and H2 (about 2.5-2.6 kb) are indicated. Invariant PgR bands are demonstrated at 9.0, 4.7, and 3.6 kb. The nucleic acid segment sequence of Figure 1 was employed as the probe.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The presently described invention provides a specific genetic diagnostic test for the identification of patients who have breast cancer as well as those persons genetically susceptible to the development thereof. The trait observed by the inventors is highly

correlatable to the incidence of breast cancer in humans. This trait is identified in particularly described DNA restriction fragments of the patient DNA, which reveal either a homozygous condition for the H2 allele or an absence of the H1 allele. Thus, H1 could be or could be revealing a tumor suppressor gene.

The method for detecting the polymorphism diagnostic for breast cancer includes a nucleic acid segment which identifies the particular genetic abnormality in a *Hin*dIII restriction fragment prepared from a DNA patient sample. Accordingly, included in the present invention disclosure is information which may be used to prepare a wide variety of nucleic acid fragments having a number of potential utilities, such as the preparation of DNA and RNA sequences in PCR and hybridization studies as probes for *in vitro* detection, as well as other useful medical and biochemical applications related to the research, diagnosis and treatment of breast cancer.

The nucleic acid fragment for the purposes of the present invention is defined as a polymer of nucleic acids. More specifically, the nucleic acid segment is defined as a polymer of nucleic acids suitable for use as a probe and sufficient to provide for the hybridization of the nucleic acid segment with segments of a patient DNA sequence which include a complementary base sequence thereto, thus identifying a *Hin*dIII RFLP characteristic of breast cancer or a susceptibility to breast cancer. Measurement of any resulting double-helix formation (hybridization) provides a relative measurement of the relatedness between the patient DNA and the nucleic acid segment of the disclosed method.

Any tissue of the patient could be used to obtain a DNA sample suitable for analysis for the herein described genetic polymorphism. By way of example, such tissues

include both tumorous and nontumorous tissues. Tissues may also be used from a variety of tissue types, such as breast tissue, tissue adjacent breast tissue, blood cells, placental tissue, thymus tissue (thymus is a lymphocyte-rich tissue important in the production and maintenance of immune cells), skin, buccal smear and hair, among others. In the most preferred embodiments of the invention, the tissue type of choice for detection of a polymorphism characteristic of breast cancer or susceptibility thereto is breast tissue or blood cells. Blood cells such as granulocytes and lymphocytes are most particularly preferred.

Even though the invention has been described with a certain degree of particularity, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art in light of the foregoing disclosure. Accordingly, it is intended that all such alternatives, modifications and variations will fall within the spirit and the scope of the invention and be embraced by the defined claims.

The following examples are presented to describe preferred embodiments in utilities of the present invention, but should not be construed as limited the claims thereof.

Example 1

SYNTHESIS OF NUCLEIC ACID SEGMENT SEQUENCES FOR IDENTIFICATION OF HUMAN BREAST CANCER AND SUSCEPTIBILITY THERETO

The present example is provided to demonstrate several preferred methods by which the nucleic acid segments for identifying human breast cancer or susceptibility thereto may be prepared. However, any of a variety of methods other than the specific methods described herein, as well as a variety of variations from

the methods specifically described herein, may be used in the practice of the present invention.

The particular methods outlined herein are the synthesis of a cDNA clone, synthesis of a sequence using PCR, and synthesis of a sequence using a synthetic oligonucleotide synthesizer.

a. cDNA probe

A DNA probe useful for the identification of a genetic polymorphism characteristic of breast cancer or susceptibility thereto in a particularly preferred embodiment is prepared using a nucleic acid sequence isolatable from a human PgR clone.

One particular PgR clone which was used in the preparation of a DNA probe suitable for use in the present invention was the human PgR clone hPR-56. This particular PgR clone, hPR-56, was isolated from a T47D-pCD library. The T47D-pCD library comprises a variety of clones containing the PgR gene sequence, and is described in Okayama et al. (1983).[32]

The particular hPR-56 clone consists of a 2.7 kb BamHI fragment of the human PgR, representing nucleotides 70-7835. The authenticity of the fragment sequence was confirmed by DNA sequence analysis.

The hPR-56 nucleic acid was digested with AccI to cleave the PgR sequence at nucleotide 1609, followed by digestion with BamHI, and a 1.23 kb AccI/BamHI fragment subcloned into pGEM (Promega, Inc.). This fragment was found by the present inventors to include a sequence useful in the identification of the genetic polymorphism for breast cancer and susceptibility to breast cancer.

b. PCR amplified probe

A synthetic oligonucleotide primer corresponding to nucleotides 1601 to 1620 of the human PgR cDNA plus an added EcoRI restriction site (5'-CGGAATTCTGCCGCAGGTCTACCCGCCC-3") and an antisense primer corresponding to nucleotides 2769 to 2788 plus an added *Hin*dIII restriction site (5'-GCAAGCTTAAGAGAAGGGGTTTCACCATC-3') can be used to amplify by polymerase chain reaction (PCR) the 1.23 kb fragment from PgR cDNA.

Following PCR-amplification, the fragment was purified by agarose gel electrophoresis and labeled with ^{32}P-dNTP by random prime labeling using a commercially available kit (Boehringer Mannheim). However, these labeling components may also be prepared individually following standard procedures (Maniatus Molecular Biology Manual).

Alternatively, a prophetic nucleic acid segment having a length of about 1.23 kb may be prepared by PCR for use in conjunction with the described methods and kits. Specifically, a nucleic acid subfragment of the 1.23 kb PgR cDNA may be prepared using two oligonucleotide sense and antisense primer pairs chosen within the 1.23 kb cDNA. These internal PgR cDNA primers should consist of 20 to 30 oligonucleotides and contain approximately 50 to 65% C/G content. The distance between the two pairs may vary, but lengths between 100 to 200 may provide optimum ease of PCR amplification and specificity for detection of the *Hin*dIII RFLP. The primers can be used to PCR amplify PgR cDNA and the amplified fragment used as a probe as outlined in Example 1. The labeled form of the fragment may be prepared using the radioisotopic label ^{32}P and used in a Southern analysis with genomic DNA from tissue specimens.

The PCR-generated probe may then also be used in hybridization reactions with the Southern nitrocellulose membranes as described in Example 2.

c. Synthetic Oligonucleotide Synthesis of a probe

The nucleic acid segment employed in the diagnostic methods and compositions herein may be synthesized as a 20 - 30 bp (base pair) synthetic oligonucleotide. As such, the oligonucleotides may be prepared via automated synthesizers, desalted, and used as probes. Most preferably, these oligonucleotides are to be labeled with a radioisotope, such as ^{32}P, using ^{32}P-dNTP's and T4 polynucleotide kinase. These labeled nucleotides may then be used in hybridization reactions with the Southern nitrocellulose membranes as described in Example 2.

Example 2
***HindIII* RFLP of Breast Cancer in Human Tissue**

The present example is provided to demonstrate the first molecular genetic evidence, through the use of RFLP analysis, that an allelic polymorphism exists which is highly correlated to the occurrence and/or subsequent development of human breast cancer.

The present example is also provided to demonstrate the utility of employing the methods described herein for the detection of particular restriction fragment length polymorphisms (RFLP) indicative or predictive of human breast cancer, as identified in digested and isolated DNA restriction fragments of a patient's DNA.

The data herein indicate an identifiable polymorphism at the H1 and/or the H2 allele is present in a statistically significant number of tissues obtained from persons with breast cancer, compared to non-tumorous human tissues obtained from persons without breast

cancer. A total of 132 breast tumors were examined for the presence of the *Hin*dIII RFLP as an indicator of breast cancer, and established that an allelic profile of H2/H2 or an allelic profile absent the H1 allele was present in tissues obtained from patients diagnosed or later diagnosed to have breast cancer.

This *Hin*dIII RFLP was found by the inventors to not display typical Mendelian distribution in the breast tumors. In the ligand-binding assays conducted by the inventors, the *Hin*dIII RFLP did not correlate with the PgR expression. This suggests to the present inventors that the RFLP is not related to the heterogeneity of PgR expression seen in breast tumors.

The present examples provide an examination of the genomic status of the PgR gene in a total of 132 breast tumor biopsies using Southern hybridization analysis.

MATERIALS AND METHODS

Cell Lines

Six human breast cancer cell lines were used in the present studies, in these particular human breast cancer cell lines were:

T47D (ATCC No.HTB 133):
ZR75 (ATCC No.CRL1500):
MDA-231 (ATCC No.HTB26):
MDA-468 (ATCC No.HTB132):
MCF-7 (ATCC No.HTB22):
MDA MB-330 (ATCC No.HTB127):

Cells were maintained as monolayer cultures in Eagles minimal essential medium with 10% fetal bovine serum. Also lines were shown to be free to microplasma contamination.

HUMAN TISSUES

Human placental tissues were obtained from a local medical center hospital. These tissues were collected immediately after parturition, trimmed free of excess connective tissue, rinsed in sterile phosphate buffered saline, frozen in liquid nitrogen, and stored at -70°C. Human breast tumor specimens, maintained at -70°C, consist of tissue remaining after routine estrogen and progesterone receptors assays and were collected from throughout the United States. Human peripheral blood leukocyte DNA was provided from local sources. Normal breasts and breast tumor pairs were obtained also from local sources and consisted of primary breast carcinoma specimens with adjacent benign breast tissue.

DNA ANALYSIS

High molecular weight genomic DNA was isolated on a Model 340A Nucleic Acid Extractor (Applied Biosystem, Inc., Foster City, CA) according to manufacturer's recommendations, and quantitated by diphenylamine assay[25]. Ten micrograms of DNA was digested with the appropriate restriction enzyme, separated by a electrophoresis on a 1% agarose gel, and transferred onto nitrocellulose by the method of Southern[20]. The nitrocellulose filters were hybridized to a ^{32}P-labelled[26] AccI/BamHI fragment of the human hPR-56 PgR cDNA clone which corresponds to the hormone binding and 3'-untranslated domains of the receptor mRNA[7] at 42°C for 16 hours in the presence of dextran sulfate. To control for variability in loading and transfer, blots were stained both before and after transfer with ethidium bromide.

HindIII POLYMORPHISM

A representative Southern hybridization analysis of DNA from breast tumors, placentas, and normal lymphocytes digested with *Hin*dIII and probed with the hPR-56 PgR cDNA clone is shown in Figure 2. Invariant bands at 9.0, 4.7,

and 3.6 kb are seen; there are also two polymorphic alleles migrating at about 2.8 and about 2.5-2.6 kb, which are labeled H1 and H2, respectively. Allele frequencies for the *Hind*III polymorphism were determined in a total of 132 breast tumors and 39 placentas (Table 1).

Table 1

DISTRIBUTION AND FREQUENCIES OF THE PR GENE *HINDIII* RFLP				
	Genotype (%)[a]			
Tissue type	H1/H1	H1/H2	H2/H2	Frequency of H1[b]
PR+ breast tumors (n-73)	67	21	12	0.77 (.71-.84)[c]
PR- breast tumors (n-59)	58	25	17	0.70(.62-.78)
placenta (n=39)	54	46	0	0.77(.67-.86)

a χ^2 analysis was performed to compare allele distribution between PR+ and PR- tumors.

b Calculations for allele frequencies based on the Hardy-Weinberg equation, giving p=0.0004 for PR+, 0.003 for PR-, and 0.06 for placenta.

c Numbers in parentheses, 95% confidence intervals.

The frequency of the H1 allele in breast tumors was 77% and 70% in PgR-positive and PgR-negative tumors, respectively, and the *Hind*III alleles were not associated with PgR expression levels in these tumors. However, the frequency of the *Hind*III RFLP in tumors did not exhibit a typical Mendelian distribution. The inventors hypothesize whether a genetic selection is occurring which accounts for the observed disequilibrium of these *Hind*III alleles in breast tumors. The placenta DNAs tested contained no detectable H2 homozygotes, although this did not reach statistical significance (p = 0.06).

Thus, a higher incidence of particular abnormalities in the form of characteristic H1 and/or H2 allelic profiles was discovered. Further, by employing the observations disclosed by the inventors, the oncogenesis of aggressive breast cancers may be even more closely examined and identified.

Example 3
H1 - Allele Free Profile
and H2/H2 Allelic Profile as a Breast Cancer Genetic Marker

The present example is provided to demonstrate that the H2 homozygotes or the H1 allele free patient profiles in the breast tumor population may represent a specific loss of the H1 allele as a marker in tumor tissue.

The *Hin*dIII RFLP was used to determine the presence of breast cancer or the susceptibility to breast cancer in lymphocyte samples. A group of 10 lymphocyte DNAs were examined: The H2 homozygote was not present in this series. These gene frequency differences reflect the three populations selected for analysis. A larger series of lymphocyte samples would determine the incidence and significance of the H2 homozygote in normal tissues.
**

While the H2/H2 genotype has been reported by others in lymphocyte DNA[8], this genotype *per se*, or the absence of the H1 genotype, has not been described as part of a method for cancer prediction or susceptibility. Note that the present inventors have found that the H1/H1 or H1/H2 genotype is not a useful predictor for this condition.

Ten breast tumors and adjacent normal breast tissue were examined according to the protocol outlined in Example 2. Five informative normal/tumor pairs heterozygous for the *Hin*dIII allele were detected, but a loss of the H1 allele in these tumors was not seen.

Thus, while it appears that not all breast tumors will show an H2/H2 genotype, the H2/H2 genotype has been shown only in individuals having a breast tumor (or at risk). A previous report has detected a loss of heterozygosity at one or more loci on the short arm of chromosome 11[10]. This allele loss was associated with grade III tumors, ER and PgR-negative tumors, and distant metastasis. However, PgR is located distal to this location at 11q21-23[8].

Polymorphisms were not detected using the restriction enzymes EcoRI and Pvu II in these same 132 tumors. PgR gene amplification (greater than 2-3 fold) was also not detected in any of the specimens using EcoRI, or Pvu II and *Hin*dIII. Alterations of 11q have been identified by banding analysis in 8/8 breast cancer cell lines. However, these alterations were highly variable in nature[24]. Others, however, have not found such a high frequency of 11q alterations in breast tumors[9], which agrees with the results obtained with this group of breast tumors. Major PgR gene rearrangement and amplification does not occur, and probably does not account for the lack of PgR expression in the majority of human breast tumors. These results suggest that based on the frequency, it appears that these reported PgR RFLP's may be useful as markers for linkage analysis.

Prophetic Example 4
Proposed Method for Predicting Patients at Risk of breast cancer with a PCR-generated nucleic acid segment

The present example is provided to demonstrate the use of the currently defined location of the RFLP for breast cancer to isolate and prepare smaller subsets of the 1.23 kb probe which recognize the *Hin*dIII RFLP. A more narrowly defined, smaller nucleic acid fragment which identified the polymorphism of a patient DNA characteristic of breast cancer or susceptibility to

breast cancer provides several practical and clinical advantages not provided by the use of a larger nucleic acid segment for DNA analysis among the practical advantages are the economical savings for preparing a smaller nucleic acid segment vs. a larger nucleic acid segment such as in the amount of nucleotides and reagents necessary for synthesis, as well as the time required to prepare, process and verify authenticity of a small verses large nucleic acid fragment. The clinical advantages associated with employing a smaller nucleic acid include a reduced risk of identifying a change in the DNA (of the patient) which perhaps overlaps the region of DNA polymorphism specific for breast cancer susceptibility, as indicative of breast cancer, but which instead reflects either a different clinical pathology, or at the very least, *not* be indicative of breast cancer or susceptibility thereto (false positive).

A smaller fragment would also be more readily utilizable in routine testing in the clinical laboratory of patient samples for testing for the presence of the RFLP described herein. Thus, the smaller fragment may be employed in a technique for the rapid diagnosing of breast cancer development, prognosis and susceptibility. On a technical level, a smaller diagnostic nucleic acid fragment sequence would facilitate a more sensitive assay for detection of the characteristic polymorphism, as "background" bands would be eliminated.

Methods

A nucleic acid fragment including a smaller fragment will be prepared as described in Example 1. The labeled form of the oligonucleotide was prepared using the radioisotopic label, ^{32}P.

By way of example, a smaller 0.48 kb nucleic acid fragment to be prepared by PCR is hypothesized.

Accordingly, this prophetic probe may be prepared by PCR. More specifically, a prophetic synthetic oligonucleotide primer corresponding to nucleotides 2301 to 2320 of the human PgR plus an added EcoRI restriction site (5′-CGGAATTCCTACAAACACGTCAGTGGGC-3′) and an antisense primer corresponding to nucleotides 2769 to 2788 plus an added *Hin*dIII restriction site (5′-GCAAGCTTAAGAGAAGGGGTTTCACCATC-3′ may be used to amplify by polymerase chain reaction (PCR) a 0.48 kb fragment of the 1.23 kb PgR. The PCR-amplified 0.48 kb fragment may then be subcloned into pGEM7zf+ and its sequence confirmed by dideoxysequence analysis.

Prophetic Example 5
Proposed Method for Predicting Patients at Risk of Breast Cancer Using PCR

The present prophetic example is provided to demonstrate the proposed use of PCR to detect the *Hin*dIII RFLP in small sample specimens. With the advent of mammography detection of small breast tumors, the amount of clinical specimen available is sometimes limiting and isolation of genomic DNA and traditional Southern analysis are not always feasible.

The inventors propose to determine the nucleotide sequence of the about 2.8 and the about 2.5-2.6 kb *Hin*dIII alleles using techniques currently available for sequence determinations in the laboratory. These sequences can be amplified from small amounts of tissue or isolated genomic DNA using an oligonucleotide primer to the 5′ end and an antisense oligonucleotide primer to the 3′ end of the 2.8 kb and/or 2.5 kb *Hin*dIII allele. The PCR products may then be digested with *Hin*dIII or directly run on a standard Southern gel and hybridized with probes recognizing the RFLPs by standard hybridization techniques. This proposed method obviates the requirement for sufficient tissue for genomic DNA

isolation and will be possible once the sequences of the 2.8 and 2.5 kb *Hin*dIII RFLPs are delineated.

The sequence of the 2.8 and 2.5 kb *Hin*dIII RFLPs are to be determined, more specifically, by obtaining a frozen normal placental tissue, obtaining genomic DNA therefrom, isolating the 2.8 kb band (H1 allele) or the 2.5 kb band (H2 allele) therefrom, preparing a subgenomic clone containing the 2.8 kb fragment or the 2.5 kb fragment, and employing the respective subgenomic clone to analyze patient sample DNA for the presence of sequences hybridizable thereto.

Preparation of a Subgenomic Clone having a 2.8 kb fragment (H1 allele) or 2.5 kb fragment (H2 allele) and Sequence Determination thereof

The placental tissue will first be analyzed for its allelic profile. A placental tissue DNA which has a discernable H1/H1 allelic profile will then be selected for use in determining a nucleotide sequence for H1 allele. The 28 kb DNA containing the H1 allele will be excised from a mock Southern Blot and the DNA cloned into a pGEM vector. The *Hin*dIII H1 allele insert will be identified by hybridization with the 1.23 kb probe, as described in Example 1. Dideoxy sequence analysis of the H1/H1 clones that hybridized with the 1.23 kb probe will then be conducted to determine the nucleotide sequence for H1.

The same procedure will be repeated with the 2.5 kb fragment obtained from placental DNA found to have an H1/H2 allelic profile. Specifically, the 2.5 kb region will be excised from a Southern Blot of DNA from placental tissue known to have an H1/H2 allelic profile.

The inventors plan to prepare a separate subgenomic clone library for H1 and H2. More specifically, a

subgenomic library out of the H2 excised region sequence and of the H1 excised region of the described placental Southern Blot analysis of DNA will be prepared. Subgenomic cloning is to be conducted according to that protocol outlined by Struhl,[32] which reference is specifically incorporated herein by reference for this purpose.

The presently proposed method will eliminate the need for tissue amounts sufficient to obtain genomic DNA isolation in a method for analyzing genetic abnormalities in a patient DNA.

BIBLIOGRAPHY

The following list represents references cited at various points in the Specification. Particular ones of these references (*), in pertinent part, are specifically incorporated herein by reference for the purposes herein indicated.

1. O'Malley BW: The steriod receptor superfamily: More excitement predicted for the future. Mol Endocrinol 4:363-369, 1990

2. Osborne CK, Yochmowitz MG, Knight WL: The value of estrogen and progesterone receptors in the treatment of breast cancer. Cancer 46:2884-2888, 1980

3. Clark GM, McGuire WL, Hubay CA, Pearson OH, Marshall JS: Progesterone receptors as a prognostic factor in stage II breast cancer. N Engl J Med 309:1343-1347, 1983

4. Jeltsch JM, Krozowski Z, Quirin-Striker C, Gronemeyer H, Simpson RJ, Garnier JM, Krust A, Jacob F, Chambon P: Cloning of the chicken progesterone receptor. Proc Natl Acad Sci USA 83:5424-5428, 1986

5. Conneely OM, Sullivan WP, Toft DO, Birnbaumer M, Cook RG, Maxwell BL, Zarucki-Schulz T, Greene GL, Schrader WT, O'Malley BW: Molecular cloning of the chicken progesterone receptor. Science 233:767-770, 1986

6. Loosfelt H, Atger M, Misrahi M, Guiochon-Mantel A, Meriel C, Logeat F, Benarous R, Milgrom E: Cloning and sequence analysis of rabbit progesterone

receptor complementary DNA. Proc Natl Acad Sci USA 83:9045-9049, 1986

7. Misrahi M, Aatger M. d'Auriol L, Loosfelt H, Meriel C, Fridlansky F, Guiochon-Mantel A, Galibert F, Milgram E: Complete amino acid sequence of the human progesterone receptor deduced from cloned cDNA. Biochem Biophys Res Commun 143:740-748, 1987

8. Law ML, Kao FT, Wei Q, Hartz JA, Greene GL, Zarucki-Schulz T, Conneely OM, Jones C, Puck TT, O'Malley BW, Horwitz KB: The progesterone receptor gene maps to human chromosome band 11q13, the site of the mammary oncogene int-2. Proc Natl Acad Sci USA 84:2877-2881, 1987

9. Trent JM: Cytogenetic and molecular biologic alterations in human breast cancer: A review. Br Cancer Res Treat 5:221-229, 1985

10. Ali IU, Lidereau R, Thjeillet C. Callahan R: Reduction to homozygosity of genes on chromosome 11 in human breast neoplasia. Science 238:185-238, 1987

11. Mackay J, Elder PA, Porteous DJ, Steel CM, Hawkins RA, Going JJ, Chetty U: Partial deletion of chromosome 11p in breast cancer correlates with size of primary tumour and oestrogen receptor level. Br J Cancer 58:710-714, 1988

12. Theillet C, Lidereau R, Escot C, Hutzell P, Burnet M, Gest J, Schlom J, Callahan R: Loss of a c-H-ras-1 allele and aggressive human primary breast carcinomas. Cancer Res 46:4776-4781, 1985

13. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL: Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235:177-182, 1987

14. Cline MJ, Battifora H, Yokota J: Proto-oncogene abnormalities in human breast cancer: Correlations with anatomic features and clinical course of disease. J Clin Oncol 5:999-1006, 1987

15. Varley JM, Swallow JE, Brammer WJ, Whittaker JL, Walker RA: Alterations to either c-erbB-2 (neu) or c-myc proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. Oncogene 1:423-430, 1987

16. Adnane J, Gaudray P, Simon M-P, Simony-Lafontaine J, Jeanteur P, Theillet C: Proto-oncogene amplication and human breast tumor phenotype. Oncogene 4:1389-1395, 1989

17. Callahan Rk, Campbell G: Mutations in human breast cancer: An overview. J Natl Cancer Inst 81:1780-1786, 1989

18. Read LD, Snider CE, Miller JS, Greene GL, Katzenellenbogen BS: Ligand-modulated regulation of progesterone receptor messenger ribonucleic acid and protein in human breast cancer cell lines. Mol Endocrinol 2:263-271, 1988

19. Wei LL, Krett NL, Francis MD, Gordon DF, Wood WM, O'Malley BW, Horwitz KB: Multiple human progesterone receptor messenger ribonucleic acids and their autoregulation by progestin agonists and antagonists in breast cancer cells. Mol Endocrinol 2:62-72, 1988

*20. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:504-517, 1975

21. Watkins PC: Restriction fragment length polymorphism (RFLP): Applications in human chromosome mapping and genetic disease research. Biotechniques 6:310-319, 1988

22. Cavenee WK, Murphree AL, Shull MM, Benedict WF, Sparkes RS, Kock E, Nordenskjold M: Prediction of familiar predisposition to retinoblastoma. N Engl J Med 314:1201-1247, 1986

23. Dressler LG, Seamer LC, Owens MA, Clark GM, McGuire WL: DNA flow cytometry and prognostic factors in 1331 frozen breast cancer specimens. Cancer 51:420-427, 1988

24. Satya-Prakash KL, Pathak S, Hsu TC, Olive M, Cailleau RE: Cytogenetic analysis on eight human breast tumor cell lines: high frequencies of 1q. 11q and HeLa-like marker chromosomes. Cancer Genet Cytogenet 3:61-73, 1981

*25. Giles KW, Myers A: An improved diphenylamine method ofor the estimation of deoxyribonucleic acid. Nature 206:93, 1965

26. Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6-13, 1983

27. Huggins *et al.* (1962) Science 137:257-262

28. Huggins *et al.* (1962) Proc Natl Acad Sci USA 48:379-386

29. Kan *et al.* (1978) Proc Natl Acad Sci USA 75:5631; Gusella *et al.* (1983) Nature 306:234 1983; D.S. Gerhard *et al.* (1984) Am J Hum Genet 36:35; M.J.M. Saraiva *et al.* (1986) Neurology 36:1413; Bartlett *et al.* (1987) Science 235:1648; P.H. St. George-Hyslop *et al.*, ibid, 885; M. Barrow *et al* (1987) Nature 326:289; J.M. Rommens *et al.* (1989) Science 245:1059; J.R. Riordan *et al.*, ibid, 1066; B. Kerem *et al.*, ibid, 1073

30. Mayslin *et al.* (1988), *Genomics, 2*:66.

31. Okayama *et al.* (1983), *Mol. Cell. Biol., 3*:280-289.

32. Struhl, Kevin (1985), *Biotechniques*, 3(6): 452-453.

SEQUENCE LISTING

I. General Information:

i) Applicant: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM

ii) Inventors: MCGUIRE, William L.
FUQUA, Suzanne A. W.

iii) Title of Invention: METHODS AND COMPOSITIONS FOR THE DETECTION AND PREDICTION OF BREAST CANCER

iv) Number of Sequences: 2

v) Correspondence Address: Arnold, White & Durkee
P. O. Box 4433
Houston, Texas 77027
USA

vi) Computer Readable Form: IBM PC-Compatible Floppy Disk,
MS-DOS, WP5.1

vii) Current Application Data: Appn. No.: Unknown
Filing Date: Unknown
Classification: Unknown

viii) Previous Application Data: Appn. No.: 07/743,606
Filing Date: 08/12/91
Classification: 435

ix) Attorney/Agent Information: Denise L. Mayfield
Reg. No. 33,732
Ref./Dkt. No. UTFK167PCT

x) Telecommunication Information: phone:(512)320-7200
fax: (512) 474-7577

2. **Information For SEQ ID NO:1:**

x) Sequence Characteristics: Length: 1.23 kb
Type: nucleic acid
Strandedness: single
Topology: linear

xi) Sequence Description:

Sequence I.D. No. 1:

10	20	30	40	50	60
GUCUACCCGC	CCUAUCUCAA	CUACCUGAGG	CCGGAUUCAG	AAGCCAGCCA	GAGCCCACAA
70	80	90	100	110	120
UACAGCUUCG	AGUCAUUACC	UCAGAAGAUU	UGUUUAAUCU	GUGGGGAUGA	AGCAUCAGGC
130	140	150	160	170	180
UGUCAUUAUG	GUGUCCUUAC	CUGUGGGAGC	UGUAAGGUCU	UCUUUAAGAG	GGCAAUGGAA
190	200	210	220	230	240
GGGCAGCACA	ACUACUUAUG	UGCUGGAAGA	AAUGACUGCA	UCGUUGAUAA	AAUCCGCAGA
250	260	270	280	290	300
AAAAACUGCC	CAGCAUGUCG	CCUUAGAAAG	UGCUGUCAGG	CUGGCAUGGU	CCUUGGAGGU
310	320	330	340	350	360
CGAAAAUUUA	AAAAGUUCAA	UAAAGUCAGA	GUUGUGAGAG	CACUGGAUGC	UGUUGCUCUC
370	380	390	400	410	420
CCACAGCCAG	UGGGCGUUCC	AAAUGAAAGC	CAAGCCCUAA	GCCAGAGAUU	CACUUUUUCA

	430	440	450	460	470	480
	CAGGUCAAG	ACAUACAGUU	GAUUCCACCA	CUGAUCAACC	UGUUAAUGAG	CAUUGAACCA
	490	500	510	520	530	540
5	GAUGUGAUCU	AUGCAGGACA	UGACAACACA	AAACCUGACA	CCUCCAGUUC	UUUGCUGACA
	550	560	570	580	590	600
	AGUCUUAAUC	AACUAGGCGA	GAGGCAACUU	CUUUCAGUAG	UCAAGUGGUC	UAAAUCAUUG
	610	620	630	640	650	660
10	CCAGGUUUUC	GAAACUUACA	UAUUGAUGAC	CAGAUAACUC	UCAUUCAGUA	UUCUUGGAUG
	670	680	690	700	710	720
	AGCUUAAUGG	UGUUUGGUCU	AGGAUGGAGA	UCCUACAAAC	ACGUCAGUGG	GCAGAUGCUG
15						
	730	740	750	760	770	780
	UAUUUUGCAC	CUGAUCUAAU	ACUAAAUGAA	CAGCGGAUGA	AAGAAUCAUC	AUUCUAUUCA
	790	800	810	820	830	840
20	UUAUGCCUUA	CCAUGUGGCA	GAUCCCACAG	GAGUUUGUCA	AGCUUCAAGU	UAGCCAAGAA
	850	860	870	880	890	900
	GAGUUCCUCU	GUAUGAAAGU	AUUGUUACUU	CUUAAUACAA	UUCCUUUGGA	AGGGCUACGA
	910	920	930	940	950	960
25	AGUCAAACCC	AGUUUGAGGA	GAUGAGGUCA	AGCUACAUUA	GAGAGCUCAU	CAAGGCAAUU
	970	980	990	1000	1010	1020
	GGUUUGAGGC	AAAAAGGAGU	UGUGUCGAGC	UCACAGCGUU	UCUAUCAACU	UACAAAACUU
30						
	1030	1040	1050	1060	1070	1080
	CUUGAUAACU	UGCAUGAUCU	UGUCAAACAA	CUUCAUCUGU	ACUGCUUGAA	UACAUUUAUC

1090	1100	1110	1120	1130	1140
CAGUCCCGGG	CACUGAGUGU	UGAAUUUCCA	GAAAUGAUGU	CUGAAGUUAU	UGCUGCACAA
1150	1160	1170	1180	1190	1200
UUACCCAAGA	UAUUGGCAGG	GAUGGUGAAA	CCCCUUCUCU	UUCAUAAAAA	GUGAAUGUCA
1210	1220	1230			
UCUUUUUCUU	UUAAAGAAUU	AAAUUUUGUG G			

2. **Information For SEQ ID NO:2:**

x) Sequence Characteristics: Length: 0.48 kb
Type: nucleic acid
Strandedness: single
Topology: linear

xi) Sequence Description:

Sequence I.D. No. 2:

10	20	30	40	50	60
CUACAAACAC	GUCAGUGGGC	AGAUGCUGUA	UUUUGCACCU	GAUCUAAUAC	UAAAUGAACA
70	80	90	100	110	120
GCGGAUGAAA	GAAUCAUCAU	UCUAUUCAUU	AUGCCUUACC	AUGUGGCAGA	UCCCACAGGA
130	140	150	160	170	180
GUUUGUCAAG	CUUCAAGUUA	GCCAAGAAGA	GUUCCUCUGU	AUGAAAGUAU	UGUUACUUCU
190	200	210	220	230	240
UAAUACAAUU	CCUUUGGAAG	GGCUACGAAG	UCAAACCCAG	UUUGAGGAGA	UGAGGUCAAG
250	260	270	280	290	300
CUACAUUAGA	GAGCUCAUCA	AGGCAAUUGG	UUUGAGGCAA	AAAGGAGUUG	UGUCGAGCUC
310	320	330	340	350	360
ACAGCGUUUC	UAUCAACUUA	CAAAACUUCU	UGAUAACUUG	CAUGAUCUUG	UCAAACAACU
370	380	390	400	410	420
UCAUCUGUAC	UGCUUGAAUA	CAUUUAUCCA	GUCCCGGGCA	CUGAGUGUUG	AAUUUCCAGA

480
UGGUGAAACC

470
UUGGCAGGGA

460
ACCCAAGAUA

450
CUGCACAAUU

440
GAAGUUAUUG

430
AAUGAUGUCU

488
CCUUCUCU

CLAIMS:

1. A method for diagnosing breast cancer, or susceptibility to breast cancer, in a patient, the method comprising:

obtaining a tissue or blood specimen from a patient;

processing the specimen to obtain DNA;

subjecting the DNA to a restriction enzyme capable of defining regions of a *Hin*dIII polymorphism, so as to provide DNA restriction fragments; and

diagnosing a patient homozygous for the H2 allele or absent a H1 allele as having breast cancer or as being susceptible to breast cancer.

2. The method of claim 1 wherein the tissue specimen is a breast tissue specimen.

3. The method of claim 1, wherein the tissue specimen comprises blood cells, buccal smear, skin, breast tissue, hair or thymus.

4. The method of claim 1, wherein the DNA comprises genomic DNA.

5. The method of claim 1 wherein the DNA restriction fragments are separated by length and probed to locate segments of an RFLP therein, said segments of the RFLP

being hybridizable to a nucleic acid segment having a H1 or H2 allele.

6. The method of claim 1 wherein a H1 allele or a H2 allele is located by probing the isolated DNA restriction fragments with a nucleic acid probe having a sequence defined in figure 1 or a fragment thereof.

7. The method of claim 1 wherein the H1 allele or the H2 allele is located with a nucleic acid probe having a sequence defined in figure 2 or a fragment thereof.

8. The method of claim 1 wherein the DNA restriction fragments are separated by electrophoresis.

9. The method of claim 1 wherein the restriction enzyme is other than EcoRI, or Pvu II.

10. The method of claim 1 wherein the restriction enzyme is *Hin*dIII restriction enzyme.

11. The method of claim 1 wherein the H1 allele is within a DNA restriction fragment having a length of about 2.8 kb.

12. The method of claim 1 wherein the H2 allele is within a DNA restriction fragment having a length of about 2.6 kb.

13. The method of claim 1 wherein the H1 allele or the H2 allele is located with a 0.48 kb nucleic acid fragment of the sequence defined in Figure 1.

14. The method of claim 1 wherein the presence of an H2 allele or an H1 allele is identified using a nucleic acid probe having a sequence defined in Figure 1, or a fragment thereof.

15. The method of claim 1 wherein the presence of an H2 allele or an H1 allele is identified with a nucleic acid probe having a sequence defined in Figure 2, or a fragment thereof.

16. The method of claim 15 wherein the nucleic acid probe is defined as a labeled 1.23 kb AccI/BamHI DNA fragment of a PgR nucleic acid sequence.

17. A method of detecting a *Hin*dIII RFLP in a DNA sample, wherein said *Hin*dIII RFLP identifies a polymorphism characteristic of breast cancer or susceptibility thereto, comprising:

treating the DNA sample with a restriction enzyme capable of producing a DNA restriction fragment having a *Hin*dIII polymorphism, to produce DNA restriction fragments;

probing the DNA restriction fragments with a nucleic acid segment capable of identifying a *Hin*dIII RFLP; and

identifying a *Hin*dIII RFLP in a segment of the separated DNA restriction fragment which hybridizes with the nucleic acid segment or a fragment thereof.

18. The method of claim 17 wherein the polymorphism for breast cancer or susceptibility to breast cancer is the absence of an H1 allele.

19. The method of claim 17 wherein the polymorphism for breast cancer or susceptibility to breast cancer is a H2/H2 allelic profile.

20. A nucleic acid segment suitable for use as a probe and capable of identifying a *Hin*dIII RFLP characteristic of human breast cancer or susceptibility thereto.

21. The nucleic acid segment of claim 20 wherein the RFLP identifies an H2/H2 gene allelic condition or the absence of an H1 gene allele.

22. The nucleic acid segment of claim 20 having a sequence defined in Figure 1 or a fragment thereof.

23. The nucleic acid segment of claim 20 having a sequence defined in figure 2 or a fragment thereof.

24. A method for diagnosing breast cancer, or susceptibility to breast cancer in a patient, the method comprising:

obtaining a tissue or blood specimen from a patient;

processing the specimen to obtain DNA;

analyzing DNA for the presence of a nucleic acid sequence hybridizable to a nucleic acid segment defined at Figure 1;

and diagnosing breast cancer or susceptibility thereto in a patient whose DNA hybridizes to the nucleic acid sequence defined at Figure 1, or a fragment thereof.

25. A kit for the prediction of breast cancer or breast cancer susceptibility in a patient, the kit comprising a hybridization probe capable of hybridizing to an H1 and H2 RFLP, wherein the probe is comprised in a suitable container.

26. The kit of claim 25, further comprising means for packaging said container.

27. The kit of claim 25, wherein the probe comprises a DNA segment.

28. The kit of claim 25, wherein the nucleic acid probe comprises a segment of about 1.23 kb or less.

29. The kit of claim 28, wherein the nucleic acid probe comprises the sequence of figure 1, or a fragment comprising a hybridizable portion of the sequence.

30. The kit of claim 28, wherein the nucleic acid probe comprises the sequence of figure 2.

31. The kit of claim 25, further comprising means for detecting hybridization between the probe and an H1 or H2 RFLP.

32. The kit of claim 31, wherein the detecting means comprises a label on the probe.

33. The kit of claim 32, wherein the label comprises an enzyme or radioactive label.

```
        10          20          30          40          50          60
GUCUACCCGC CCUAUCUCAA CUACCUGAGG CCGGAUUCAG AAGCCAGCCA GAGCCCACAA

        70          80          90         100         110         120
UACAGCUUCG AGUCAUUACC UCAGAAGAUU UGUUUAAUCU GUGGGGAUGA AGCAUCAGGC

       130         140         150         160         170         180
UGUCAUUAUG GUGUCCUUAC CUGUGGGAGC UGUAAGGUCU UCUUUAAGAG GGCAAUGGAA

       190         200         210         220         230         240
GGGCAGCACA ACUACUUAUG UGCUGGAAGA AAUGACUGCA UCGUUGAUAA AAUCCGCAGA

       250         260         270         280         290         300
AAAAACUGCC CAGCAUGUCG CCUUAGAAAG UGCUGUCAGG CUGGCAUGGU CCUUGGAGGU

       310         320         330         340         350         360
CGAAAAUUUA AAAAGUUCAA UAAAGUCAGA GUUGUGAGAG CACUGGAUGC UGUUGCUCUC
```

FIG. 1A

370 380 390 400 410 420

CCACAGCCAG UGGGCGUUCC AAAUGAAAGC CAAGCCCUAA GCCAGAGAUU CACUUUUUCA

430 440 450 460 470 480

CCAGGUCAAG ACAUACAGUU GAUUCCACCA CUGAUCAACC UGUUAAUGAG CAUUGAACCA

490 500 510 520 530 540

GAUGUGAUCU AUGCAGGACA UGACAACACA AAACCUGACA CCUCCAGUUC UUUGCUGACA

550 560 570 580 590 600

AGUCUUAAUC AACUAGGCGA GAGGCAACUU CUUUCAGUAG UCAAGUGGUC UAAAUCAUUG

610 620 630 640 650 660

CCAGGUUUUC GAAACUUACA UAUUGAUGAC CAGAUAACUC UCAUUCAGUA UUCUUGGAUG

670 680 690 700 710 720

AGCUUAAUGG UGUUUGGUCU AGGAUGGAGA UCCUACAAAC ACGUCAGUGG GCAGAUGCUG

FIG. 1B

```
      730        740        750        760        770        780
UAUUUUGCAC CUGAUCUAAU ACUAAAUGAA CAGCGGAUGA AAGAAUCAUC AUUCUAUUCA

      790        800        810        820        830        840
UUAUGCCUUA CCAUGUGGCA GAUCCCACAG GAGUUUGUCA AGCUUCAAGU UAGCCAAGAA

      850        860        870        880        890        900
GAGUUCCUCU GUAUGAAAGU AUUGUUACUU CUUAAUACAA UUCCUUUGGA AGGGCUACGA

      910        920        930        940        950        960
AGUCAAACCC AGUUUGAGGA GAUGAGGUCA AGCUACAUUA GAGAGCUCAU CAAGGCAAUU

      970        980        990       1000       1010       1020
GGUUUGAGGC AAAAAGGAGU UGUGUCGAGC UCACAGCGUU UCUAUCAACU UACAAAACUU
```

FIG. 1C

```
      1030        1040        1050        1060        1070        1080
CUUGAUAACU UGCAUGAUCU UGUCAAACAA CUUCAUCUGU ACUGCUUGAA UACAUUUAUC

      1090        1100        1110        1120        1130        1140
CAGUCCCGGG CACUGAGUGU UGAAUUUCCA GAAAUGAUGU CUGAAGUUAU UGCUGCACAA

      1150        1160        1170        1180        1190        1200
UUACCCAAGA UAUUGGCAGG GAUGGUGAAA CCCCUUCUCU UUCAUAAAAA GUGAAUGUCA

      1210        1220        1230
UCUUUUUCUU UUAAAGAAUU AAAUUUUGUG G
```

FIG. 1D

```
        10          20          30          40          50          60
CUACAAACAC GUCAGUGGGC AGAUGCUGUA UUUUGCACCU GAUCUAAUAC UAAAUGAACA

        70          80          90         100         110         120
GCGGAUGAAA GAAUCAUCAU UCUAUUCAUU AUGCCUUACC AUGUGGCAGA UCCCACAGGA

       130         140         150         160         170         180
GUUUGUCAAG CUUCAAGUUA GCCAAGAAGA GUUCCUCUGU AUGAAAGUAU UGUUACUUCU

       190         200         210         220         230         240
UAAUACAAUU CCUUUGGAAG GGCUACGAAG UCAAACCCAG UUUGAGGAGA UGAGGUCAAG

       250         260         270         280         290         300
CUACAUUAGA GAGCUCAUCA AGGCAAUUGG UUUGAGGCAA AAAGGAGUUG UGUCGAGCUC

       310         320         330         340         350         360
ACAGCGUUUC UAUCAACUUA CAAAACUUCU UGAUAACUUG CAUGAUCUUG UCAAACAACU
```

FIG. 2A

```
                370       380       390       400       410       420
UCAUCUGUAC UGCUUGAAUA CAUUUAUCCA GUCCCGGGCA CUGAGUGUUG AAUUUCCAGA

                430       440       450       460       470       480
AAUGAUGUCU GAAGUUAUUG CUGCACAAUU ACCCAAGAUA UUGGCAGGGA UGGUGAAACC

                490       500       510       520       530
CCUUCUCUUU CAUAAAAAGU GAAUGUCAUC UUUUUCUUUU AAAGAAUUAA AUUUUGUGG
```

FIG. 2B

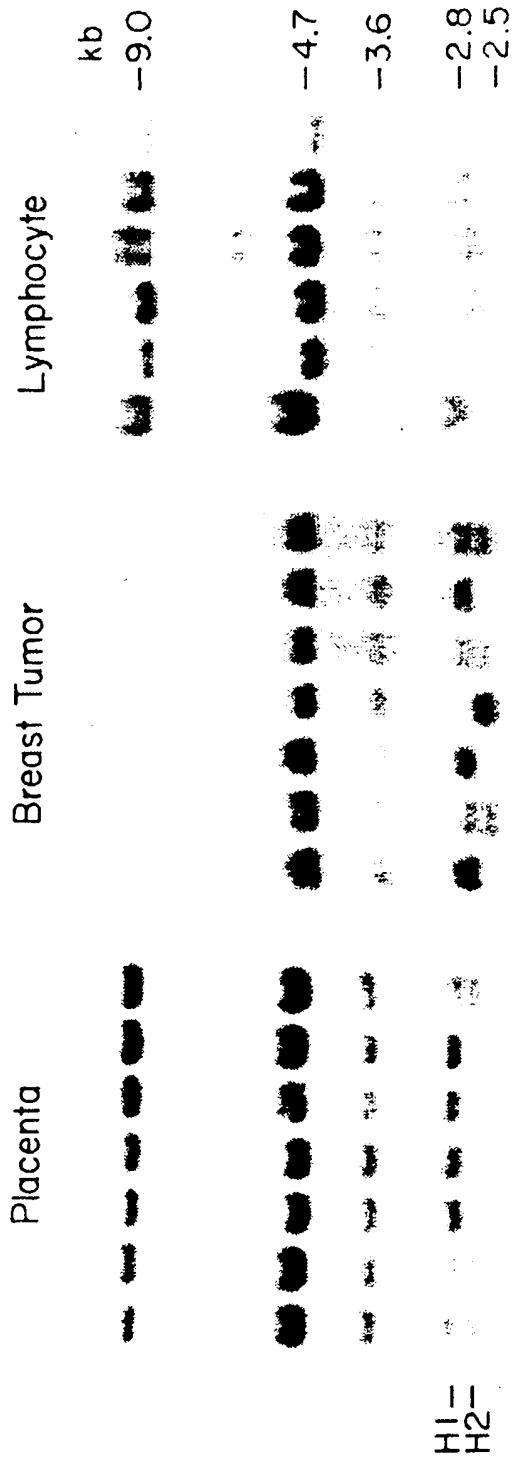


FIG. 3

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 92/06701

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12Q1/68

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X	SCIENCE vol. 238, 9 October 1987, LANCASTER, PA US pages 185 - 188 I.U. ALI ET AL. 'Reduction to homozygosity of genes on chromosome 11 in human breast neoplasia' cited in the application see the whole document, especially the abstract, fig. 1 and table 1 --- -/--	1,10

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
04 NOVEMBER 1992	23.11.92
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	LUZZATTO E.R.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS. vol. 143, no. 2, 1987, DULUTH, MINNESOTA US pages 740 - 748 MISRAHI M. ET AL. 'Complete amino acid sequence of the human progesterone receptor deduced from cloned DNA' cited in the application see the whole document, especially the abstract	20-24
X	CANCER RESEARCH vol. 48, no. 14, July 1988, PHILADELPHIA,USA pages 4045 - 4048 J.S. LEE ET AL. 'Multiple restriction fragment length polymorphism of the human epidermal growth factor receptor gene' see abstract see "Discussion"	1,10
A	SURGICAL FORUM vol. 38, 1987, CHICAGO,USA pages 402 - 405 A.-M. MARCOUX ET AL. 'Restriction fragment length polymorphisms of type I collagen as genetic markers for breast cancer' see the whole document	1,9
A	BREAST CANCER RESEARCH AND TREATMENT vol. 14, no. 1, 1989, THE HAGUE,THE NETHERLANDS pages 57 - 64 F.F.PARL ET AL. 'Genomic DNA analysis of the estrogen receptor gene in breast cancer' see abstract	1,9
A	FR,A,2 628 441 (IMPERIAL CHEMICAL INDUSTRIES) 15 September 1989 see page 2, line 6 - page 3, line 25; claims	1

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9206701 SA 63706

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 04/11/92
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
FR-A-2628441	15-09-89	GB-A- 2217450	25-10-89
		JP-A- 2009398	12-01-90